

Combining multiple data sources in functional genomics for improving genome-wide inferences

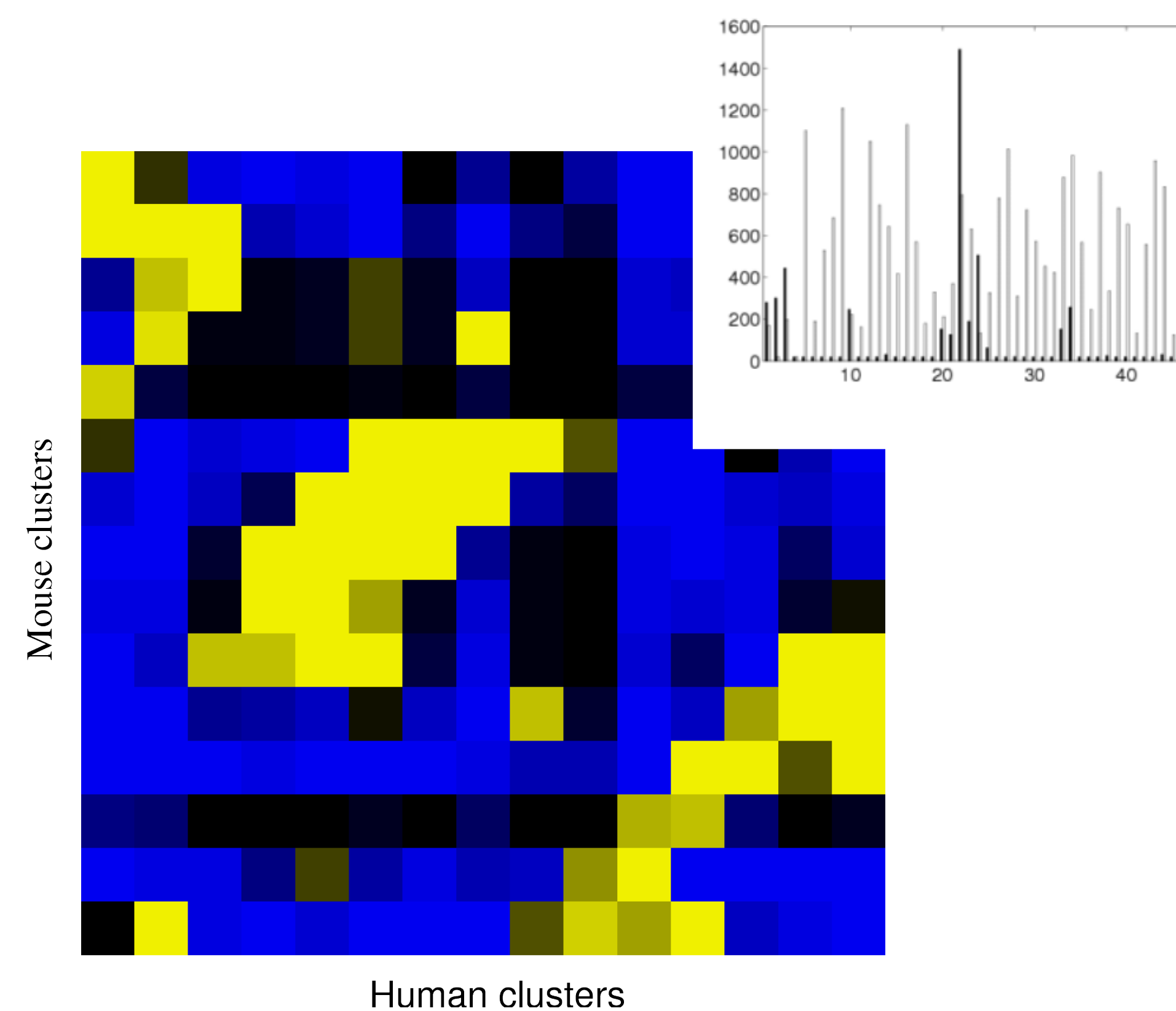


Leo Lahti, Samuel Myllykangas, Sakari Knuutila, Jaakko Hollmén, Eero Castrén, and Samuel Kaski



Data analysis tools

associative clustering is one of the developed data fusion techniques; it is used for exploring dependencies between data sources

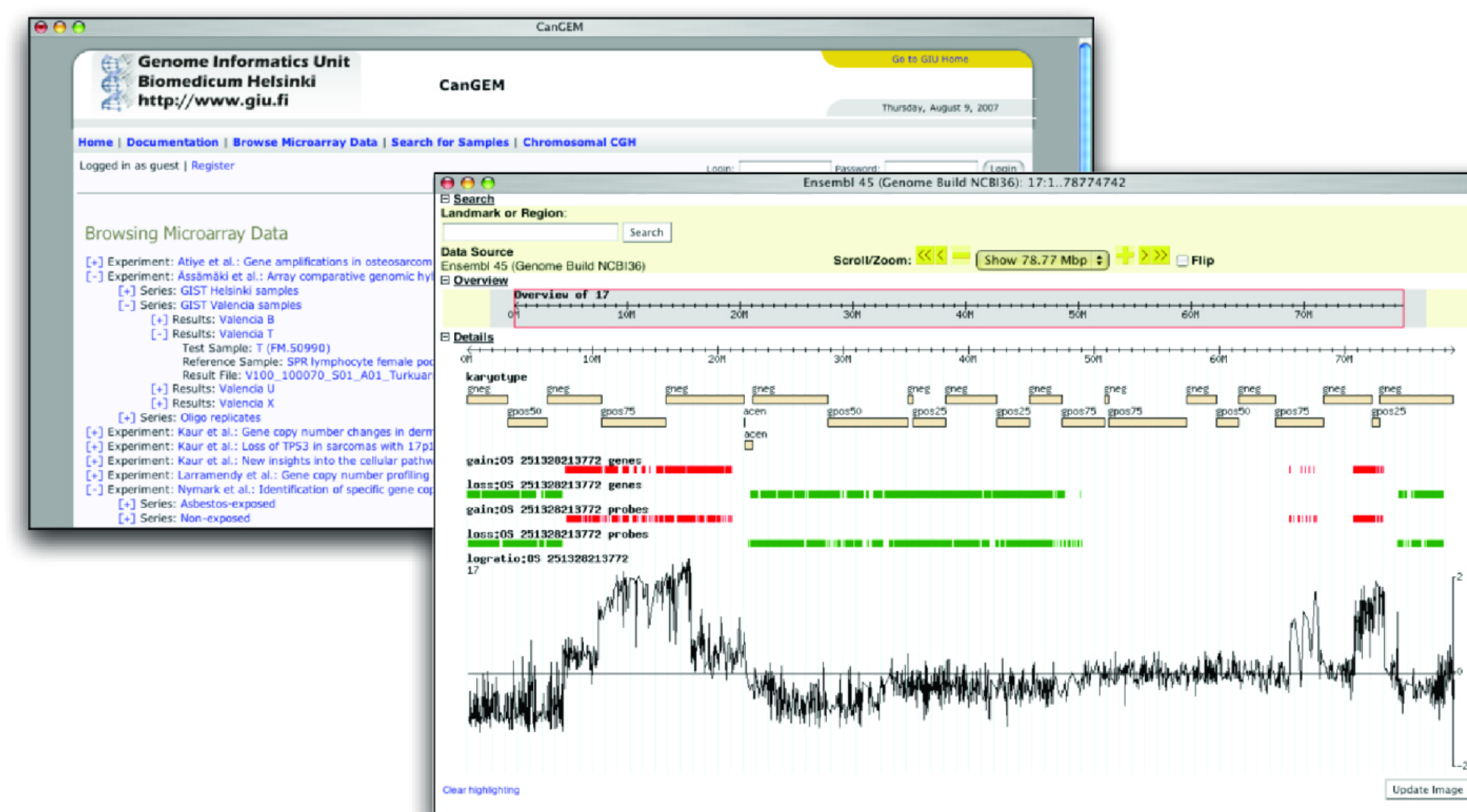


Overview of the project

Our key research question is how to take advantage of existing, partially representative data sets of different types to support inferences in biological and medical questions. We have developed general data analysis methods for this task. Developed and existing methodologies have been applied in cancer research and comparative functional genomics. Here we show representative examples of the results.

Collecting representative data

publicly available CanGEM database facilitates further mining of gene copy number and other large-scale genomic data



Biomedical applications

genome-wide integration strategies highlight relationships between various cancer types. Integration of gene copy number and expression data has led to identification of putative target genes that could have value in clinical applications

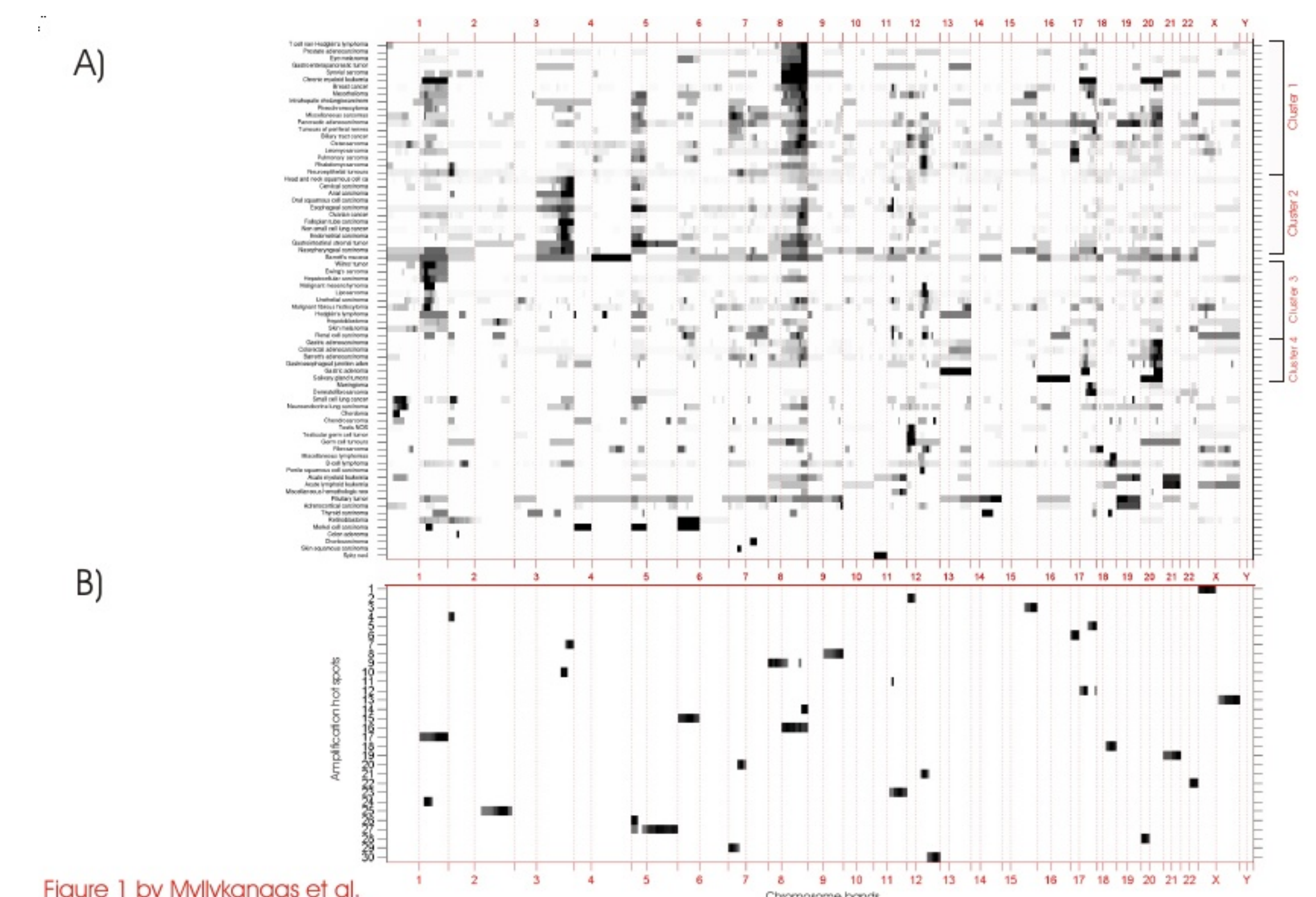


Figure 1 by Myllykangas et al.

Members of the consortium

Consortium leader: Samuel Kaski, professor, Lab. of Computer and Information Science, TKK

Other project leaders of the consortium:
Jaakko Hollmén, chief research scientist, Dpt. of Computer Science and Engineering, TKK
Eero Castrén, professor, Neuroscience Center, University of Helsinki
Sakari Knuutila, professor, Depts. of Pathology and Medical Genetics, University of Helsinki

Doctoral students of the consortium:

Anne Tyybäkinöja	Leo Lahti
Arto Klami	Merja Oja
Hanna Vauhkonen	Mikko Korpela
Ilari Scheinin	Pamela Lindholm
Ioana Borze	Penny Nymark
Janne Toivola	Salla Ruosaari
Jarkko Venna	Samuel Myllykangas
Juha E.A. Knuutila	Suvi Savola

Other researchers of the consortium:

Jaakko Peltonen, researcher, TKK
Janne Nikkilä, researcher, TKK
Sami Khuri, visiting professor, San Jose State University

References

Myllykangas S, Tikka J, Böhling T, Knuutila S, Hollmén J. Classification of human cancers based on DNA copy number amplification patterns. Submitted 10/2007.

Scheinin I, Myllykangas S, Borze I, Knuutila S, and Saharinen J. CanGEM: mining gene copy number changes in cancer. Nucl Acids Res. In press.

Nymark P, Lindholm PM, Korpela MV, Lahti L, Ruosaari S, Kaski S, Hollmén J, Anttila S, Kinnula VL, and Knuutila S. Gene expression profiles in asbestos-exposed epithelial and mesothelial lung cell lines. BMC Genomics 8:62, 2007.

Sairanen M, O'Leary OF, Knuutila JE and Castrén E. Chronic antidepressant treatment selectively increases expression of plasticity-related proteins in the hippocampus and medial prefrontal cortex of the rat. Neuroscience 144: 368-374, 2007.

Kaski S, Nikkilä J, Sinkkonen J, Lahti L, Knuutila J, and Roos C. Associative clustering for exploring dependencies between functional genomics data sets. IEEE/ACM Transactions on Computational Biology and Bioinformatics. Special Issue on Machine Learning for Bioinformatics – Part 2, vol. 2, nr. 3, pp. 203-216, 2005.

For more information, see
<http://www.cis.hut.fi/projects/mi/sysbio/>